Linguistic Barriers to Financial Flows

Undergraduate Research Thesis

Presented in Partial Fulfillment of the Requirements for graduation "with Honors Research Distinction in Economics" in the undergraduate colleges of The Ohio State University

by
Annette Gailliot

The Ohio State University
April 2018

Project Advisors:
Professor Bruce Weinberg, Department of Economics
Dr. Ethan Doetsch, Department of Economics

Abstract
Historical ties and linguistic similarities between countries are highly correlated with outcomes in the financial sector. Various types of linguistic relationships all affect information flows between populations, and countries with similar languages can more easily share ideas and are likely to have come from similar historical groups. There is evidence that linguistic factors have a strong relationship with bilateral trade, and I test whether this is also true for financial flows. Using bilateral data on the debt and equity flows between the 35 OECD countries in 2002 to 2012, I examine the changing patterns of financial flows and language between countries. I find that the relationship between linguistic similarity and equity flows is significant and large in magnitude before 2008. After 2008, this pattern holds. However, the relationship between linguistic similarity and debt flows is not strong before or after 2008. Despite integration under the European Union and institutions like the OECD, historical linguistic patterns of financial flows still hold in times of economic prosperity.

Keywords: international finance, linguistics, financial flows

Acknowledgements: Thank you to Dr. Ethan Doetsch, Dr. Bruce Weinberg, and Dr. Amanda Robinson for all the generous input and guidance throughout developing this research project. All errors are my own.

I. Introduction

Despite globalization and market integration under the European Union (EU), countries send more capital to linguistically similar countries than dissimilar ones. Financial theory is still exploring why international diversification in Equity portfolios is not the norm, when the benefits from diversification are widely known (French and Poterba, 1991). A similar relationship between language and international trade has been explored in depth (Melitz and Toubal, 2012-2017), but financial flows and trade flows follow different patterns (Nardo et al, 2017). The exact mechanisms through which language affects international trade and financial flows are still under dispute, and it is unclear whether linguistic differences act as a simple transaction cost (Melitz et al, 2008) or as a proxy for ethnic similarity and trust (Alesina et al, 2003).

Using novel linguistic and financial datasets, I analyze the relationship between linguistics and financial flows. As new datasets on bilateral financial flows become available, we can learn more about the patterns of financial trade. Hobza and Zeugner's "FinFlows" dataset provides a wide breadth of information by estimating the Debt and Equity flows between 80 countries, but it is also incredibly specific by providing bilateral information to and from each of these countries[1]. In addition to this, Melitz and Toubal have recently compiled a dataset of linguistic similarity between countries[2]. They use this information to show that language is a complex variable that should not necessarily be simplified into a binary shared official language variable, as has frequently been done in the past (Ginsburgh and Weber, 2011). These datasets allow for an in-depth analysis of the patterns between linguistic ties and financial flows.

In this study, I attempt to shed light on the relationships between language and financial flows. I do this by examining various linguistic measures. Since the OECD countries all have well-developed financial systems, as well as a wide variety of linguistic differences, I analyze the

[1] This is initially provided in Hobza and Zeugner, 2014 and is further described in their web appendix: http://www.zeugner.eu/studies/finflows/. They have also recently updated the dataset, as detailed in Nardo et al, 2017. This paper also describes an analysis they perform using the FinFlows dataset. As the 2017 data has not yet been published, I use the 2014 dataset.

[2] Melitz and Toubal compiled this dataset throughout 2012-2017, and their published paper in 2014 analyzes international trade using this dataset. Further details from the CEPII can be found here: http://www.cepii.fr/cepii/en/bdd_modele/presentation.asp?id=19

35 OECD countries[3] provided in Zeugner and Hobza's 2014 version of "FinFlows." This also allows me to focus on developed countries rather than the different financial markets that exist in emerging and developing countries. By analyzing the period from 2002 to 2012, I am also able to gain some insight into the changing financial and linguistic patterns that occur during and after the 2008 financial crisis. These changing patterns before and after the financial crisis provide an intriguing opportunity to analyze what factors may be the most important to financial flows in different broad economic conditions.

While these analyses do not provide causal evidence of language affecting financial flows, they do highlight a significant, large in magnitude, and persistent relationship between linguistic variables and financial flows. At first glance, in an age of economic integration and globalization, it is surprising that language could have such a strong relationship with financial flows. Despite EU integration, these linguistic patterns still exist, at least when the overall economy is strong. Since one of the OECD's goals is to "restore confidence in markets and the institutions that make them function[4]," it is interesting to explore how these markets reacted to the uncertainty of the 2008 financial and following debt crisis. The results of this study suggest that a deeper, causal analysis could yield valuable results about the nature of linguistic effects on financial flows, especially during times of financial crisis.

The rest of this paper is organized as follows. Section 2 briefly explores the literature surrounding language's relationships with international financial and trade flows. Section 3 details the main data sources I use. Section 4 provides the methods of analysis. Section 5 analyzes and interprets my results. Section 6 discusses limitations. Section 7 concludes.

II. Literature

This research operates in the intersection of international trade and international finance. Within these two related fields, language's effects have been widely studied, and many avenues through which language affects economic outcomes have been identified. The main ones

---

[3] Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, South Korea, Latvia, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States.

[4] http://www.oecd.org/about/

discussed include language as: a way to measure trust and ethnic ties between linguistic populations (Guiso et al, 2006; Tabellini, 2010; Desmet et al, 2011) and as a transaction cost and barrier to trade (Melitz, 2008). Although much research argues for the importance of language, it is still not clear how to proxy and define a concept as broad as language. Some use language as a binary indicator of whether or not countries have the same official language (Ginsburgh and Weber, 2011), and others combine language with other variables such as within ethnolinguistic fractionalization (Alesina et al, 2003; Desmet et al, 2016). However, Nardo, et al (2014) have found that the effects of language on international trade are underestimated on the order of at lease one-half when language is measured through official language rather than more complex measures (Nardo, et al, 2014).

To explore the mechanisms of language, Melitz and Toubal (2014) use five different measures of linguistic similarity. These measures describe the commonalities between the language populations in two countries, and they discuss how each identifies a different avenue of communication. The five measures are common official language (COL), common spoken language (CSL), common native language (CNL), linguistic similarity based on the Ethnologue and Fearon and Laitin measures (Prox1), and linguistic similarity (Prox2) based on the Automated Similarity Judgment Program (ASJP). The value of using multiple linguistic measures lies in identifying mechanisms through which language is used. By using COL and CSL to capture the transaction cost/ease of communication mechanism of language and CNL, Prox1, and Prox2 to capture the historical ties mechanism of language, Melitz and Toubal provide a unique way to analyze language's relationship with economic outcomes.

Melitz (2008) describes COL to measure the presence of an overhead cost to communication. If two countries share an official language, then there is no additional communication cost to the transaction (i.e. translation). The people in these countries are already capable of sharing messages without incurring any additional private cost and so COL can be used to measure the "transaction cost" mechanism of shared language. Melitz and Toubal posit that CSL measures the direct ease of communication while CNL isolates the ethnicity and trust mechanisms of language, since native language measures the language taught at birth and thus likely the language of the parents. The linguistic proximity measures, Prox1 and Prox2, further explore the ethnic/historical ties inherent in the development of language by looking at the differences in the native languages themselves. By isolating the various mechanisms of

language, these measures provide an interesting way to understand how differences between populations affect economic outcomes. Language is also an interesting measure that can capture how financial interactions vary at different levels. Households engage in finance differently than large firms, and the mechanisms of language may capture these differences. With global firms, the historical or ethnic ties that language captures may not be as important. These firms may act the same regardless of where specific firms are located since the global headquarters may govern business decisions. However, at the household level, the transaction costs of trying to engage in a different language are likely much higher since households engage in relatively fewer financial transactions than firms do. Additionally, households may be much more likely to value the trust that a historical or ethnic tie captures. Compared to large firms that engage in many financial transactions and have market power, households have less information to make financial decisions. Therefore, historical or ethnic ties may be more important to them as they attempt to invest compared to how firms make investment decisions. Language is a useful way to measure multiple mechanisms important in making financial decisions and language can capture how aspects of these decisions change at different levels.

While the precise way to measure language is still debated, the relationship between language and trade have become accepted such that it is now common to control for language in analyses of bilateral trade (Melitz and Toubal 2012-17). Although the geography of financial flows themselves can differ significantly from trade flow patterns (Nardo et al, 2014), language may have similar effects in both international trade and financial debt and equity flows. These effects of language have been widely studied within the international finance literature, specifically within the context of the Equity Home Bias puzzle. This ongoing puzzle highlights the fact that individuals and institutions disproportionately form equity portfolios of investments from their home country and rarely invest in foreign equity markets (French and Poterba, 1991; Werner and Tessar, 1995). While the advantages of diversification in equity portfolios are widely known, there is still little diversification in equity portfolios through international investment[5]. Theories of exchange rate risk (Adler and Dumas, 1983), capital immobility (Black, 1974), and information asymmetry (Merton, 1987) have all been analyzed as potential

[5] Finance theory argues that investors in the United States can optimize their portfolios through holding 60% domestic and 40% international (Shapiro 1999), and yet United States investors are actually holding 92.3% domestically (French and Poterba, 1991).

explanations behind this puzzle, with information asymmetry receiving the most attention and support in recent years.

The information asymmetry hypothesis posits that incomplete information between foreign and domestic markets leads to an investment bias toward domestic markets and away from foreign markets (Gehring, 1993). This broad concept of information asymmetry can take shape in multiple ways and has been analyzed in terms of the influence that culture and language can have on financial transactions. For example, it has been documented that Finnish investors are more likely to hold, buy, and sell stocks with firms that share similar languages and cultures (Grinblatt and Keloharju, 2001). Furthermore, cultural institutes that promote familiarity with cultures and languages have positive economic effects on trade and FDI through increasing trust and reducing transaction costs (Lien and Lo, 2017). The specific cases studied in both of these papers highlight the mechanisms through which language affect financial interactions.

This can also be studied in a broader context of international bilateral financial flows. Studies have analyzed how institutions affect bilateral financial flows (Papaioannou, 2008) and how information frictions affect investments in portfolio equity and debt securities (Daude et al, 2008). Within these analyses, language is often used as simply a binary control variable – 1 if two have countries have the same official language and 0 otherwise. Using the expanded linguistic data of Melitz and Toubal, I perform analyses of the relationships between financial flows and linguistic similarity. Language is an interesting measure to analyze since it can capture some of the differences in people that could affect financial interactions. Something must explain the Equity Home Bias' persistence throughout nearly the last thirty year, and it is possible that cultural and ethnic differences could help us understand this irrational behavior still seen in financial markets. Using new measures of linguistic similarity provide a way to capture these difficult to quantify differences.

III. Data

This project is unique in the use of two novel datasets. For language, I use the Centre d'Études Prospectives et d'Informations Internationales language dataset, as constructed and analyzed by Melitz and Toubal (2012-17), and for financial flows, I use Hobza and Zeugner's relatively new "FinFlows" dataset on the financial flows of countries in the European Union (2014). Using these two datasets, I combine financial and linguistic data on the 35 OECD

countries from 2002-2012[6]. This allows me to analyze the patterns of financial flows in developed countries before and after the financial crisis. Table 1 reports the summary statistics.

Language:

While these measures of linguistic similarity are useful, it can be difficult to understand how they are different and how they are related to specific linguistic mechanisms. In Table A of the Data Appendix, I describe examples of the linguistic measures for the United States, Spain, the United Kingdom, and Turkey. COL is a simple binary measure where, in this example, the USA and the UK have a 1 for COL and Turkey and the UK would have a 0. CSL and CNL both capture percentages of populations. The USA and Spain have (0.96)*(0.27) = 0.2592 for the shared populations of English speakers and (0.16)*(0.99) = 0.1584 for shared populations of English speakers. Overall, the USA and Spain then have a percentage of their populations that both speak the same language equal to 0.2592+0.1584 = 0.4176. This means that 41.76% of the populations in the USA and Spain speak the same language. However, in these samples, there are portions of the population that speak more than one language. This leads the CSL measure in some instances to be more than 100%. To account for this, Melitz and Toubal adjust the final CSL measure by $CSL = \max(\alpha) + (\alpha - \max(\alpha))(1 - \max(\alpha))$ where α is the initial raw CSL measure for each language. Therefore, the final CSL measure between the USA and Spain is actually 38%. CNL is calculated similarly, but the effects of double-counting are negligible since few people in the OECD countries are taught multiple languages at birth. For Turkey and Spain, CNL would be zero, but between the USA and the UK, CNL would be 0.82*0.92 = 0.7544. This means that none of the populations in Turkey and Spain were taught the same language at birth but that 75% of the populations in the USA and the UK are taught the same language at birth.

The linguistic proximity measures, Prox1 and Prox2, are based off of the CNL measure. These measures take into account the proportion of populations sharing the same native language and then compare the languages themselves. In COL, CSL, and CNL, the measurements are

[6] For Belgium and Luxembourg, the linguistic data was combined into a "BLX" variable that covered both Belgium and Luxembourg. However, there were bilateral financial flows data for both Belgium and Luxembourg. To be able to use these flows data for these two countries, I used the combined "BLX" linguistic data for both Belgium and Luxembourg individually.

based only on the differences in populations. The linguistic proximity measures take this into account, but they also look at the differences in the language itself. Prox1 does this by comparing the family trees of the languages whereas Prox2 compares the cognates of the same words in the languages. The language differences between Spain and Portugal illustrate how these measures are different. Spain's native language is Spanish whereas Portugal's native language is Portuguese. Spain and Portugal have a Prox1 measure of 0.75, illustrating how Spanish and Portuguese are close together in the West Iberian Romance linguistic family trees. However, the cognates in these languages are different and so the Prox2 measurement is a much lower 0.42. All of these measures range from zero to 1, where COL is binary and the other four measures are percentages.

Melitz and Toubal describe their specific methodology further in their paper. While these mechanisms of linguistic similarity between groups may intuitively make sense, the five variables are highly correlated and likely capture some overlapping effects. Table 2 reports the correlation matrix. Recognizing this, it is important to take care in interpreting how specifically identified these mechanisms of language are.

Financial Flows:

To study international financial flows, Hobza and Zeugner (2014) have compiled a database of bilateral financial flows between European Union countries and their major world partners. Ranging from the years 2002-2012, they use this data to analyze financial flow patterns in response to the financial crisis of 2008 and the debt financing occurring through 2009-11. In their analysis, they find that trade flows follow markedly different patterns than financial flows. Hobza and Zeugner initially use stock Balance of Payments statistics and then convert them into flows. From these flows, I focus on the two main measures - Debt and Equity, as well as their sum, labeled "Value". Positive values represent investment into a country while negative are disinvestment. Each observation indicates a bilateral Equity or Debt flow in real Euros (bn) from the "reporting" country to the "partner" country. Debt includes portfolio debt, other investments, and the debt flows through the European Central Bank. Equity combines portfolio equity and foreign direct investment. They exclude financial derivatives from their analysis, since bilateral information is rarely available. They also account for additional flows through the European Central Bank. This is key in analyzing financial flows in 2010-2012 as Greece, Portugal,

Ireland, and Spain received financial assistance from other OECD countries through the European Central Bank. The instruments and assets that make up these measures are included as an appendix, and a detailed description can be found in Hobza and Zeugner's online data appendix[7]. This bilateral data allows for a specific analysis on financial trading partners.

Controls:

In any cross-country analysis there are a wide array of omitted variables. By using the 35 OECD countries, I can analyze well-developed financial markets. However, there are other potential complicating factors discussed in the literature that I attempt to control for. I include country specific fixed effects for both the reporting and partner country. This means that for every regression I include dummy variables specific to the country sending the capital and for the country receiving the capital. Including country specific dummies controls for anything specific to the country that is present in every observation of that country. Additionally, for each observation, I control for reporting and partner country GDP per capita (IMF). This takes into account the overall size of the economy and makes the outcomes comparable between countries with large and small economies. I also control for whether the countries share a legal origin. In this dataset, there are four legal origins (British, French, Scandinavian, and German). The literature discusses how countries with British legal origin are more likely to have creditor protections that encourage financial growth than countries with other legal origins (La Porta et al, 2008). I include legal origin to account for this as well as to account for broader historical factors that legal origin may proxy. In some of the literature (Beck et al, 2002), legal origin is used to control for historical ties instead of language, but in my analysis I control for legal origin and instead analyze the various mechanisms of language through multiple linguistic variables. I also control for the geographical distance between countries (Mayer et al, 2011), whether both countries are in the European continent, and whether the countries share a border. This controls for some of the increased transaction costs that further geographical distance may add that can affect financial interactions between countries. Lastly, I include the log of the reporting and partner country's populations (OECD, 2018). I provide summary statistics and a correlation matrix for all controls in Tables 1 and 2, respectively.

[7] http://www.zeugner.eu/studies/finflows/

IV. Methods

To isolate the various relationships between linguistic similarity and financial flows, I utilize multiple OLS regressions. In each test, I perform two-way clustered standard errors, clustering on the reporting and partner country following the method Spolaore and Wacziarg (2009) use in a similar bilateral analysis and presented in Cameron et al, 2011. Two-way clustering controls for standard errors present within groups but random across groups. The two groups I cluster on in this analysis are the reporting and partner countries. Within the reporting or partner countries, there could be systematic error that is unrelated to the other variables I include. For example, there could be some reporting error in Germany's Debt flows data that occurs in every year. However, Germany's reporting error would not be correlated with error reported by France. Clustering in this way increases the standard errors and confidence intervals, leading to more accurate significance results.

In the following discussion, *Y* is one of the financial flow variables (Debt, Equity, or Value)[8], *LinguisticSimilarity* identifies the language variables (COL, CSL, CNL, Prox1, or Prox2), *i* is the reporting country, *j* is the partner country, and $\boldsymbol{\lambda}$ is the vector of controls described above. I account for time in various ways, which will be described below. In the following analyses, I test each language variable individually[9]. By using five variables associated with language and testing them separately from the others, I provide a robustness check of the relationship between language and financial flows.

Pooled Cross-Section with Time Trend

To first study the relationship between language and financial flows, I analyze a pooled cross-section with a time trend to observe the overall relationship. I use the following equation:

$$Y_{ij} = \beta_0 + \beta_1 LinguisticSimilarity_{ij} + \beta_2 Year + \boldsymbol{\beta_3 \lambda}_{ij} + \varepsilon_{ij}$$

where Year is the time trend dummy for each year, reported for each linguistic variable. The results are reported in Table 3 and are significant and large in magnitude for Equity. However, it

[8] I do not take logarithms of the financial flows variables because they are directional, with positive values representing investment and negative values representing disinvestment.
[9] I also test multiple language variables together in an attempt to isolate the mechanisms through which language acts. This was not fruitful, but I discuss it in the limitations section.

seems there were significant flow changes in specific years rather than an overall trend (see Figure 1). To further analyze this, I perform additional tests where I specify individually on year rather than an overall time trend.

Bilaterally by Year

One way to analyze the different relationships between language and financial flows over time is to look at each year individually. In this analysis, I analyze the data bilaterally by year in accordance with the following equation:

$$Y_{ij,t} = \beta_0 + \beta_1 LinguisticSimilarity_{ij} + \boldsymbol{\beta_2 \lambda_{ij,t}} + \varepsilon_{ij,t}$$

where t is each year from 2002 to 2012. By looking at each year individually, I test how the language and financial flow relationships changes over time. The results are reported in Table 4. These results have a wide variety of outcomes, but there seems to be some difference over the years. The coefficients of the relationship between the linguistic variables and Equity seems to have a more significant and larger relationship in the years following 2008, while their relationship with Debt is minimal in most of the years. Looking at the overall relationships with Value, it seems that there are only a few years where linguistic similarity has a strong relationship with Value flows.

Pooled Pre- and Post-2008 with Time Trend

To further inspect these relationships before and after 2008, I perform a similar pooled cross-section with time trend, but I separate the time periods into Pre-2008 and Post-2008. The equation for this is as follows:

$$Y_{ij} = \beta_0 + \beta_1 LinguisticSimilarity_{ij} + \beta_2 Year + \beta_3 Pre2008 + \boldsymbol{\beta_4 \lambda_{ij}} + \varepsilon_{ij}$$

where Pre2008 separates the tests into the years 2002-2007 and 2008-2012. Year is the time trend dummy for each year in Pre- or Post-2008 periods. The results are reported in Tables 5 and 6. This indicates that Equity has a strong and large in magnitude relationship with linguistic variables prior to 2008 as well as post-2008. Debt flows do not seem to be significantly related to linguistic similarity, but the significance does increase minimally following 2008.

V. Analysis & Discussion

There is a significant relationship between the linguistic variables and Equity flows. Even when controlling for a variety of potential confounding variables, these results still hold. In particular, this pattern holds for all linguistic similarity measures even in terms of border countries. Between border countries, linguistically similar countries are likely to send more capital to one another than linguistically dissimilar ones (see Figure 2). For example, 99% of the population in Austria and Germany speak the same language compared to the 36% that speak the same language in Austria and Hungary. Austria's average Equity flows to Germany are nearly twice the size of Austria's Equity flows to Hungary, relative to GDP. Even though Germany and Hungary both physically border Austria, there is a large difference in the amount of Equity flows between border countries.

In the overall pooled cross-section with time trend (see Table 3), linguistic variables have a general positive trend with financial flows. The linguistic variables' coefficients corresponding to Equity tend to be highly significant and large in magnitude. If two countries share an official language, the amount of Equity traded between the two countries is 1,836 million Euros more than two countries that do not share an official language, all else equal. This corresponds to more than a third of a standard deviation increase in Equity flows. These results hold for Equity at the 1% level. However, Debt flows are not significantly related to any of the linguistic variables. Value is significant with COL, CNL, Prox1, and Prox2 but only at the 10% level. While Equity flows are significantly related to linguistic similarity, these results do not hold for Debt flows when analyzing the overall time period.

Analyzing results bilaterally by year somewhat supports these overall trends (see Table 4), but not all variables broken up by year are significant. Within Equity, the significance of the linguistic variables is mixed, especially in 2003-2006 and 2008-2010. However, the strength and magnitudes in 2002, 2007, 2011, and 2012 resemble the previous pooled cross-section trends. It is interesting that the relationship between linguistic similarity and Equity flows is strongest in 2007. This indicates that Equity flow patterns may change leading up to financial crises and that linguistic similarity is capturing some of this change. Whether this means that it was simply cheaper to invest in linguistically similar countries or that countries were only investing in countries they had historical ties with and thus were more likely to trust is unclear. However, there could be interesting Equity flow patterns directly leading up to the financial crisis. The Debt variables show little correlation with any of the linguistic variables. Equity's strength in

2007 comes through in the Value results. However, most of the significance in the other years disappears.

Pooling the Pre- and Post-2008 periods with separate time trends sheds light on the changing patterns between linguistic similarity and financial flows. Generally, the linguistic variables are positively related to Equity flows prior to 2008 (see Table 5). After 2008, linguistic variables are only related to Equity (see Table 6), with minor significance on some of the linguistic variables' relationship with Debt flows. Prior to 2008, countries with populations sharing 100% of native languages have almost two-thirds of a standard deviation more in Equity flows than two countries sharing no native languages. This is significant at the 10% level. After 2008, this relationship holds with Equity, increasing in significance and slightly increasing in magnitude. Two countries with languages sharing 100% of the same cognates have over half a standard deviation more Equity flows than two countries sharing 0% of the same cognates. This is significant at the 5% level. However, two countries with populations speaking 100% of the same languages share over a quarter of a standard deviation fewer Debt flows than two countries that do not speak the same languages. This is significant at the 5% level, and only one other linguistic similarity measure is significant with Debt flows after 2008. However, this may be due to the drying up of Debt flows after 2008. The overall Debt flows change from 15.96 trillion before 2008 to 8 trillion after while the total Debt flows disinvested increase from 2.02 trillion before 2008 and 7.08 trillion after 2008 (see Figure 3). Similar changes occurred in Equity as well, with overall investment decreasing and total disinvestment increasing. It could be that the changing relationships between Debt flows and linguistic similarity may be capturing the overall decrease in total Debt investment and increase in disinvestment rather than a changing importance of linguistic similarity.

The patterns between linguistic similarity and financial flows are seen both in the levels and slopes of the relationships. Looking at the graphs in Figure 4, the general slopes of the Debt-language relationships change direction before (dashed lines) and after (dotted lines) 2008. However, the intercepts stay the same or decrease. There is only a slight level shift in Debt before and after 2008, but there is a large change in the slopes of the Debt-language relationships. This change in slopes likely indicates the changing environment and potential omitted variables that could be affecting financial flows. The slope levels off, but, as described above, most of the Debt-language patterns are not statistically significant. In Equity, the

relationship stays largely the same, as can be seen in the general slopes of the dashed and dotted lines. The intercepts stay the same or slightly shift up, which depicts the changing baseline of financial flows between countries if they have no linguistic similarity, holding all else equal. Linguistic similarity with Equity is significant before and after 2008, but the magnitudes slightly increase or stay the same. This is demonstrated by the small increase in the intercepts of the Equity-language relationships after 2008. The Value-language relationship intercepts tend to decrease and the slopes somewhat level off after 2008. Value is showing the combination of the Debt and Equity relationships with language. Even when analyzing these relationships with their respective 95% confidence intervals, these patterns all hold.

The general positive relationship between linguistic similarity and financial flows is present before and after 2008. However, this Equity relationship changes slightly in magnitude after 2008. For the linguistic coefficients that are significant in relation to Debt after 2008, the positive pattern disappears, but the trends of the relationship between linguistic similarity and Debt after 2008 are highly variable. This indicates that linguistic similarity is not the only variable related to financial flows. Although Debt is somewhat negatively related to linguistic relationships following 2008, these results are not robust to all linguistic similarity measure. However, even when controlling for time trends, country specific fixed effects, GDP per capita, legal origin, geographic distance, border effects, European continent effects, population effects, and two-way clustering of the standard errors, these results hold.

The 2008 Financial Crisis & European Debt Financing

One potential explanation for the change in the relationships between the linguistic relationships and Debt flows before and after 2008 is the Debt financing that occurred following the 2008 financial crisis. Prior to the financial crisis, countries' financial flows were not following similar patterns as countries' linguistic relationships. After the financial crisis of 2008, the coefficients of language's relationship with Debt flows change. For this period, linguistic similarity was somewhat negatively related to debt flows. In their 2017 paper using the FinFlows dataset, Nardo et al analyze this pattern of Debt flows between countries in the European Union. They found that countries with a capital account surplus, 'core' countries, financed countries with a capital account deficit, 'periphery countries,' through Debt financing. They find that the current account balance of the Euro area never exceeded 1% because a 'core'

group of countries (primarily Germany and the Netherlands) with current account surpluses of nearly 6% of GDP financed the 'periphery' group of countries (Greece, Ireland, Italy, Portugal, and Spain) with collective current account deficits of nearly 7% of their combined GDP. The leveling off of the slopes in the relationship between Debt flows and linguistic similarity after 2008 is potentially capturing this changing behavior in Debt financing following the financial crisis. However, even when I performed all of the previous tests while dropping all observations including one of the 'core' or 'periphery' countries, these patterns still persisted. Therefore the changing linguistic patterns aren't due entirely to outliers in the 'core' and 'periphery' countries and are then also occurring in some of the countries that weren't as instrumental in debt financing throughout the European Union.

VI. Limitations

While this evaluation provides convincing evidence of a relationship between linguistic similarity and Equity flows, there are some limitations. The main limitation is that I am not capturing any causal relationship. While a causal relationship may exist between linguistic relationships and financial flows, my tests do not include an identification strategy to isolate these effects. Despite this, these results are still strong, large, and useful in understanding the relationships between language and financial flows. By examining these patterns, we can better create a causal model to examine the most salient features of the relationships. This is something I plan to explore in further research by better identifying the pathways through which language affects financial interactions. Specifically, it could be interesting to perform these same analyses but specify the level at which the financing occurs. Linguistic similarity may be more important in financial decisions at the household level as opposed to the firm or country level. Since firms and the country flows are likely much larger than household flows, much of the patterns in this analysis could be driven by the firm or country participants. Additionally, the interesting mechanisms that language captures may not be present in certain firms. In global firms, even though the reported flows may be in a country like Germany, the firm headquarters may be in the United States. Then, the linguistic similarity that would be driving business decisions in my model should be due to the United States and not Germany. My model does not capture the level or financial interaction or the global relationships particular firms may have. This would hide the relationship between household financial flows and linguistic similarity. If linguistic

similarity matters more at the household level, then testing at this level could be a useful way to explore the mechanisms of language. However, this data is not readily available in the FinFlows dataset. In order to specify the level of flows, I would have to look at the source data used in FinFlows or find a different dataset containing this information. While the question may be interesting, the availability of data specific enough may pose a problem.

Using the five different language variables separately acts as a robustness check to language measures and their potential mechanisms. However, it is also possible that I am capturing reverse causality or that language proxies for other variables I have not included. Reverse causality seems unlikely upon further analysis. If this is what I am capturing, it would mean that financial flows are driving the linguistic patterns. This would imply a deep, historical trend of debt and equity flows leading populations to integrate and spread their languages. This seems unlikely since linguistic development comes before financial development and since language is generally thought of as necessary to engage in financial transactions. Reverse causality in this analysis does not hold much weight. However, omitted variables could be a larger issue. I attempt to control for this through my vector of control variables, but there are omitted variables in a cross-country analysis. Since language is potentially capturing the changing Debt financing after 2008, it is possible that language is capturing something else before 2008. I am unable to identify the specific mechanism that is driving these patterns, but, regardless of this, the relationship is still strong. This indicates that there could be interesting underlying relationships present in a deeper, causal analysis.

Along with this, language's relationship with financial flows could be overstated since I do not specify how much the English language drives this relationship. English is sometimes considered the "language of business," and it is possible that much of the pattern I found between linguistic similarity and financial flows before 2008 comes from communication in English. However, this issue does not necessarily disprove my hypothesis. English similarity is still linguistic similarity, and part of my analysis is also trying to identify if language is capturing "trust" between countries. Even if English is the language of business, that wouldn't necessarily mean English is the majority language learned at birth in a country. If English is not the majority native language but business is done in English, then the native language measure may still capture other factors that play a part in engaging in business (i.e. historical ties). Language in this analysis may capture more than the potential transaction costs associated with not using

English. It would be interesting to explore English's effects in particular in future research, but that is not the aim of this project.

Within the methods of this study, I wanted to develop a stronger econometric model. However, since the language similarities are constant over time, I could not perform a fixed effects panel analysis. This would have been useful to observe changes in the financial flows over time. Since this requires differencing, it would have made the linguistic variables meaningless. Instead, I looked at the patterns specific to each year as well as including a time trend. Additionally, separating the relationships into Pre- and Post-2008 allowed for some analysis of time specific factors. By analyzing these correlations in different time periods, I've been able to capture some of these time related patterns without using a fixed effects model. I also plan to include country fixed effects.

VII. Conclusion

Despite the OECD and EU promoting economic and cultural integration, language still has a strong relationship with equity flows. The transaction costs of communicating through different languages may not be the only way language affects finance. Historical ties and trust between populations may also affect finance on a global scale. Even when analyzing multiple measures of linguistics, I find that linguistic similarity between populations has a strong relationship with financial flows. However, language may be related to debt and equity flows in different ways. Linguistic similarity and equity flows share a strong, positive relationship before and after the financial crisis of 2008. This could indicate that equity home biases are strong, regardless of the overall economic environment. Debt flows, on the other hand, seem to be more responsive to economic conditions. Before the financial crisis, linguistic similarity and debt flows are not related, but this relationship gains some significance after 2008. As Europe used debt financing in response to the financial crisis, the importance of linguistic relationships may have changed. In an uncertain economy, the historical ties and lowered transactions costs associated with linguistic similarity may be more significantly related to debt flows. Observing these patterns is useful in understanding how populations engage in finance. Even while the ease of communication between countries increases due to organizations like the EU and the OECD, language still matters in finance.

References

Adler M. & Dumas B. (1983). "International Portfolio Choice and Corporation Finance: A Synthesis. " *The Journal of Finance, 38(3),* 925-984.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). "Fractionalization." *Journal of Economic Growth, 8,* 155-194.

Beck, T. and Demirgüç-Kunt, A. & Levine, R. (2002). "Law and Finance: Why Does Legal Origin Matter?" *World Bank Policy Research Paper No. 2904*.

Black, F. (1974). "International capital market equilibrium with Investment barriers." *Journal of Financial Economics, 1,* 337-352.

Cameron, C.A., Gelbach, J., & Miller, D.L. (2011). "Robust Inference With Multiway Clustering."*Journal of Business & Economic Statistics: 29(2),* 238-249.

Daude, C. & Fratzscher, Ma. (2008). "The pecking order of cross-border investment." *Journal of International Economics, Elsevier, 74(1),* 94-119.

Desmet, K., Gomes, J. & Ortuno-Ortin, I. (2016). "The Geography of Linguistic Diversity and the Provision of Public Goods." *CESifo Working Paper Series, 6,238*.

Desmet, K., Ortuño-Ortín, I. & Wacziarg, R. (2011). "The Political Economy of Linguistic Cleavages." *VIVES Research Centre for Regional Economics.*

French, K & Poterba, J. (1991). "Investor Diversification and International Equity Markets." *American Economic Review, 81,* 222-226.

Gehring, T. (1993). "An information Based Explanation of the Domestic Bias in International Equity Investment." *Scandinavian Journal of Economics*, 95(1), 97-109.

Ginsburgh, V., & Weber, S. (2011). "How Many Languages Do We Need?: The Economics of Linguistic Diversity." *Princeton University Press.*

Grinblatt, M & Keloharju, M. (2001). "How Distance, Language, and Culture Influence Stockholdings and Trades." *The Journal of Finance, 56(3),* 1053-1074.

Guiso, L., Paola S. & Zingales, L. (2006). "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives, American Economic Association, 20(2),* 23-48.

Hobza, A. & Zeugner, S. (2014). "Current accounts and financial flows in the Euro Area." *Journal of International Money and Finance, 48,* 291-313.

International Monetary Fund World Economic Outlook Database: *GDP per capita*. http://www.imf.org/ external/ns/cs.aspx?id=28.

La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2008). "The Economic Consequences of Legal Origins." *Journal of Economic Literature, 46(2),* 285-332.

Lien, D. & Lo, M. (2017). "Economic Impacts of Cultural Institutes." *The Quarterly Review of Economics and Finance, Elsevier, 64(C),* 12-21.

Mayer, T. & Zignago, S. (2011). "Notes on CEPII's distances measures: the GeoDist Database." *CEPII Working Paper 2011-25.*

Melitz, J. & Toubal, F. (2012). "Native language, spoken language, translation, and trade." *CEPII, Working Papers 2012-17.*

Melitz, J. & Toubal, F. (2014). "Native Language, Spoken Language, Translation and Trade." *Journal of International Economics, 92(2),* 351-363.

Marc J. & Ottaviano, G.I. (2008). "Market Size, Trade, and Productivity." *Review of Economic Studies, 75(1),* 295-316.

Merton R. (1987). "An equilibrium Market Model with Incomplete Information." *Journal of Finance, 42(3),* 483-510.

Nardo, M., Ndacyayisenga, N., Pagano, A. & Zeugner, S. (2017). "Finflows: a database

for bilateral financial investment stocks and flows." *European Commission JRC Technical Reports.*

OECD (2018), Population (indicator). doi: 10.1787/d434f82b-en (Accessed on 05 April 2018).

Papaioannou, E. & Siourounis, G. (2008). "Democratisation and Growth." *The Economic Journal, 118,* 1520-1551.

Shapiro, A.C. (1999). "Multinational Financial Management." *New York, New York: John Wiley & Sons.*

Spolaore, E. & Wacziarg, R. (2009). "The Diffusion of Development." *The Quarterly Journal of Economics, MIT Press, 124(2),* 469-529.

Tabellini, G. (2010). "Culture and Institutions: Economic Development in the Regions of Europe." *Journal of the European Economic Association 8 (4),* 677–716.

Tesar, L.L. & Werner, I.M. (1995). "Home Bias and High Turnover." *Journal of International Money and Finance, 14(4),* 467-492.

Data Appendix:

A. Language

Country	*Official*	*Spoken*	*Native*
USA	English	English 0.96, Spanish 0.16	English 0.82, Spanish 0.15
Spain	Spanish	Spanish .99, English .27, French .12	Spanish 0.89
UK	English	English 0.99, French 0.23, German 0.09, Spanish 0.08	English 0.92
Turkey	Turkish	Turkish .99, English .17	Turkish .93

B. Financial Flows

Instrument	*Definition & Components*
Equity	This measure of Equity includes portfolio Equity and foreign direct investment. *Portfolio Equity:* Equity comprises all instruments and records acknowledging claims on the residual Value of a corporation or quasi-corporation, after the claims of all creditors have been met, including: - Shares and stocks, participation documents, depository receipts and shares in mutual funds or investment trusts are included. *Foreign Direct Investment (FDI):* Investment of more than 10% ownership of a firm. Investments of less than 10% ownership are considered *portfolio* Debt or Equity. - Debt: marketable securities (e.g. bonds and non-participating preference shares), loans, deposits, Debt securities, trade credit, and other account receivable/payable - Equity: concerns the shareholder's funds. It contains acquisition or disposal of Equity capital, revaluations which are not distributed as dividends. It includes items such as common and preferred shares, reserves and dividends. The Equity flows include reinvestment earnings.
Debt	This measure of Debt includes Portfolio Debt, Other Investments, and the TARGET2 Balance Estimate. *Portfolio Debt:* Debt instruments are instruments that require the payment of principal and/or interest at some point(s) in the future, including: - *Long-term Debt:* convertible bonds into Equity, nonparticipating preferred stocks, zero-coupon and other deep-discounted bonds, indexed bonds, asset backed securities - *Short-term Debt:* treasury bills, bankers' acceptances, certificates of deposit, commercial paper *Other Investments:* A residual category that includes positions and transactions other than FDI, portfolio investment, financial derivatives, employee stocks options, and reserve assets. - Currency and deposits; loans; nonlife insurance technical reserves, life insurance and annuities entitlements, pension entitlements, and provisions for calls under standardized guarantees; trade credits and advances; SDR allocations *TARGET2 Balance Estimate:* Measure of net asset holdings through the European Central Bank.

Graphics Appendix:

Figure 1: Financial Flows by Year and Transaction Type

This graph details the financial flows of Debt, Equity, and Value over the years 2002-2012 in real Euros (bn) for all 35 OECD countries combined. The box-and-whisker plots are shown, with the median, 75th, and 25th percentiles indicated.

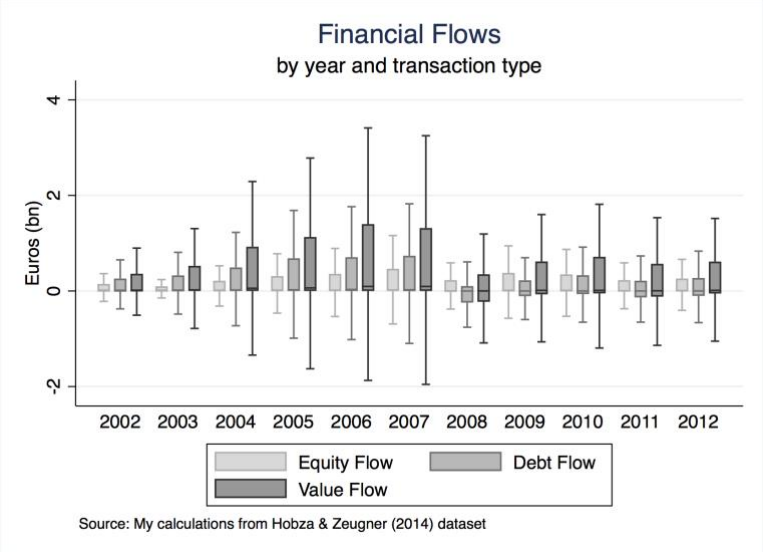


Figure 2: Average Value Flows 2002-2012 between Border Countries by Linguistic Similarity

This graph plots the linguistic similarity between bordering countries for the various common linguistic variables (excluding common official language since it is binary) versus average overall Value flows between bordering countries for the years 2002-2012. This shows the pattern that countries tend to send more capital to the countries they border when those countries are linguistically similar. I plotted this same relationship for Debt and Equity in various time periods and similar results occurred (those graphs can be provided upon request).

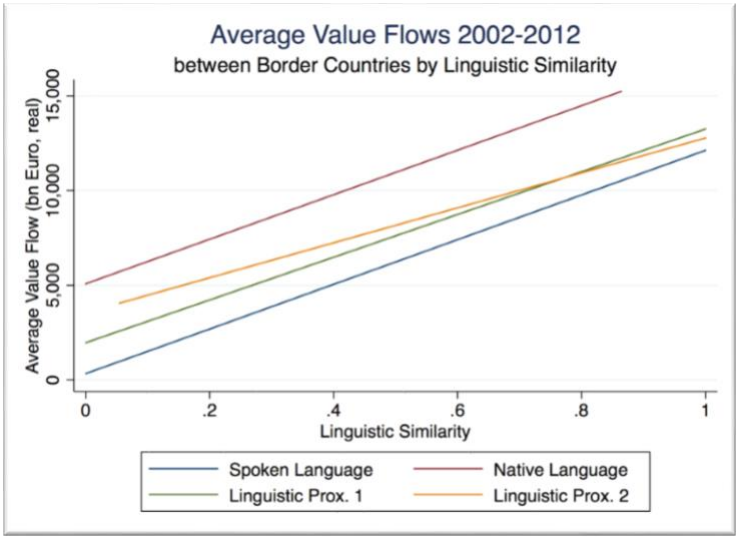
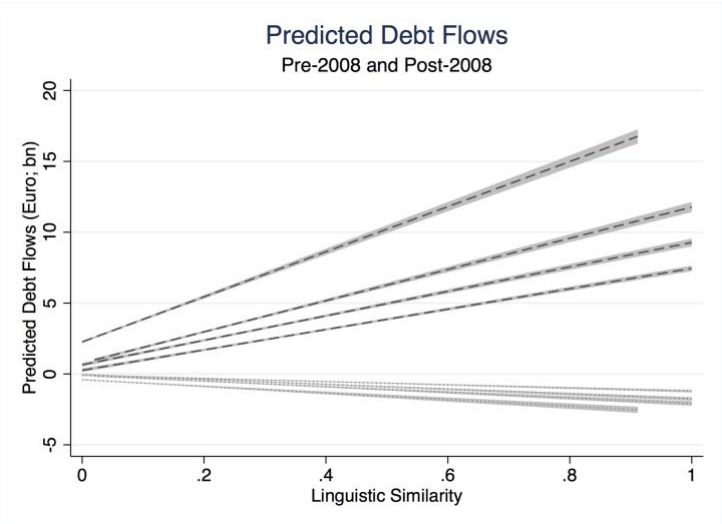


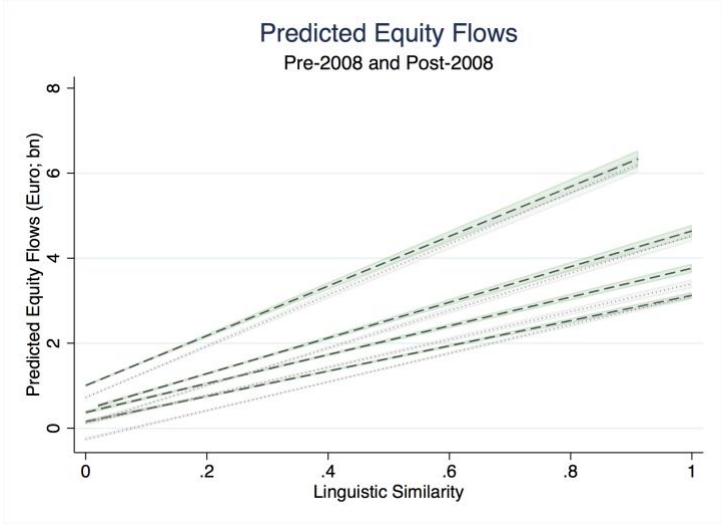
Figure 3:
Average Positive and Negative Flows Before and After 2008

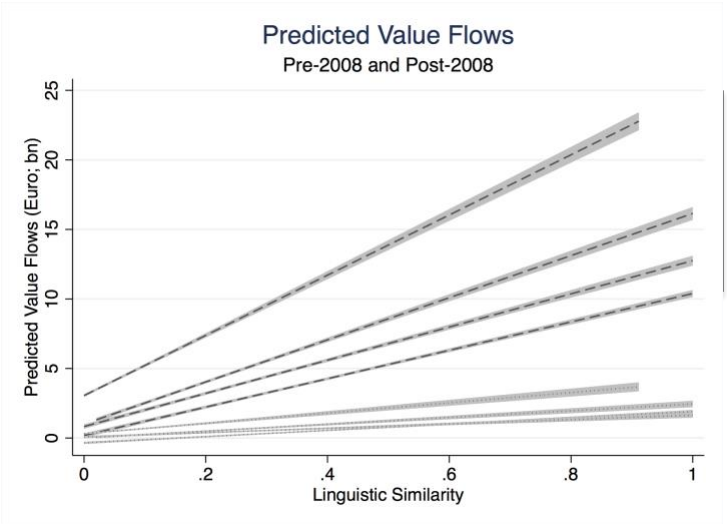
Positive and Negative Flow Sums	**Equity**	**Debt**
Pre-2008 Investment (+)	4.95	15.96
Standard Deviation	(0.005)	(0.012)
Post-2008 Investment (+)	2.57	8
Standard Deviation	(0.005)	(0.008)
Pre-2008 Disinvestment (-)	0.9	2.02
Standard Deviation	(0.003)	(0.004)
Post-2008 Disinvestment (-)	1.97	7.08
Standard Deviation	(0.005)	(0.009)

Figure 4: Graphs of Pooled Pre- and Post-2008 with Time Trend Regressions

These graphs depict the regression analyses done in the pooled Pre- and Post-2008 with time trend regressions. The dashed lines depict Pre-2008 patterns, and the dotted lines depict Post-2008 patterns. All lines are presented with their respective 95% confidence intervals in light gray. The confidence intervals for Equity Pre-2008 are in light green. Common official language is excluded from these graphics since it is a binary variable. All linguistic variables presented here are percentages of linguistic similarity from 0-100%.







Legend:

Pre-2008 — — — —

Post-2008 ·····

Tables Appendix:

Table 1: Mean and variations
Equity, Debt, and Value in millions of Euros (real, relative to 2012)

Variables	*N*	Mean	SD	Minimum	Maximum
Common Official Language	13,068	.074	.262	0	1
Common Spoken Language	13,068	.347	.268	0	.9998
Common Native Language	13,068	.031	.134	0	.911
Linguistic Proximity (Ethnologue)	12,276	.260	.229	0	1
Linguistic Proximity (ASJP)	12,276	.193	.186	.020	1
Equity	13,090	825	4,915	-86,224	112,155
Debt	13,090	1,139	9,515	-181,546	229,567
Value	13,090	1,934	11,417	-165,019	260,520
Reporter GDP	13,090	34,022	21,073	3,660	115,762
Partner GDP	13,090	34,022	21,073	3,660	115,762
Common Legal Origin	13,090	0.255	0.436	0	1
In Europe Dummy	13,090	0.546	0.498	0	1
Border Countries Dummy	13,090	0.063	0.243	0	1
Geographic Distance	13,090	5,363	5,249	161	19,540
Log Reporter Population	13,090	16.363	1.534	12.569	19.565
Log Partner Population	13,090	16.363	1.534	12.569	19.565

Table 2. Correlations

	COL	CSL	CNL	Prox1	Prox2	Value	Equity	Debt	CLO	GDP_R	GDP_P	EUR	Border	Dist.	Log Pop_R	Log Pop_P
COL	1.00															
CSL	0.47	1.00														
CNL	0.74	0.45	1.00													
Prox1	0.49	0.66	0.58	1.00												
Prox2	0.63	0.66	0.77	0.91	1.00											
Value	0.14	0.15	0.16	0.14	0.15	1.00										
Equity	0.18	0.17	0.17	0.16	0.16	0.55	1.00									
Debt	0.08	0.10	0.11	0.09	0.10	0.91	0.19	1.00								
CLO	0.32	0.09	0.34	0.25	0.36	0.05	0.06	0.03	1.00							
GDP_R	0.13	0.36	0.06	0.22	0.17	0.10	0.12	0.06	-0.08	1.00						
GDP_P	0.13	0.36	0.06	0.22	0.17	0.07	0.12	0.03	-0.08	0.06	1.00					
EUR	-0.09	0.20	-0.11	0.01	-0.04	0.02	0.00	0.02	0.07	0.09	0.09	1.00				
Border	0.22	0.17	0.17	0.17	0.20	0.11	0.12	0.07	0.29	0.03	0.03	0.19	1.00			
Dist.	0.95	-0.12	0.10	0.05	0.06	-0.05	-0.05	-0.04	-0.09	-0.09	-0.09	-0.82	-0.24	1.00		
Log Pop_R	0.04	-0.17	0.09	-0.001	0.02	0.09	0.09	0.07	0.03	-0.21	0.01	-0.23	0.04	0.13	1.00	
Log Pop_P	0.04	-0.17	0.09	-0.001	0.02	0.11	0.06	0.10	0.03	0.01	-0.21	-0.23	0.05	-0.13	-0.04	1.00

Table 3. Pooled Cross-Section with Time Trend

VARIABLES	Equity					Debt					Value				
Common Official Language	1,836***					543.0					2,241*				
	(3.071)					(0.526)					(1.774)				
Common Spoken Language		2,023**					496.9					2,418			
		(2.140)					(0.544)					(1.606)			
Common Native Language			3,321**					3,485					6,814*		
			(2.428)					(1.257)					(1.786)		
Linguistic Proximity (Ethnologue)				2,126**					2,039					4,092*	
				(2.483)					(1.288)					(1.803)	
Linguistic Proximity (ASJP)					2,455**					2,199					4,633*
					(2.447)					(1.192)					(1.743)
Year Time Trend	9.788	9.788	9.788	7.298	7.298	-179.9	-179.9	-179.9	-196.5	-196.5	-204.2	-204.2	-204.2	-224.5	-224.5
	(0.244)	(0.244)	(0.244)	(0.175)	(0.175)	(-1.433)	(-1.433)	(-1.433)	(-1.469)	(-1.469)	(-1.447)	(-1.447)	(-1.447)	(-1.500)	(-1.500)
Reporter GDP per capita	0.00919	0.00919	0.00919	0.0101	0.0101	0.00457	0.00457	0.00457	0.00471	0.00471	0.0136	0.0136	0.0136	0.0144	0.0144
	(0.737)	(0.736)	(0.721)	(0.779)	(0.764)	(0.0910)	(0.0909)	(0.0909)	(0.0900)	(0.0901)	(0.231)	(0.231)	(0.231)	(0.237)	(0.237)
Partner GDP per capita	-0.0177	-0.0177	-0.0177	-0.0176	-0.0176	0.0653	0.0653	0.0653	0.0687	0.0687	0.0736	0.0736	0.0736	0.0783	0.0783
	(-0.807)	(-0.804)	(-0.810)	(-0.779)	(-0.779)	(1.356)	(1.355)	(1.354)	(1.367)	(1.366)	(1.527)	(1.538)	(1.530)	(1.569)	(1.569)
Common Legal Origin	98.60	223.9	104.1	103.3	-4.362	451.7	501.1	177.6	209.3	132.8	559.5	718.7	262.1	307.5	116.1
	(0.431)	(0.873)	(0.710)	(0.463)	(-0.0233)	(1.000)	(0.930)	(0.476)	(0.473)	(0.317)	(0.946)	(1.009)	(0.580)	(0.550)	(0.230)
In Europe ==1	-77.06	238.7	324.3	617.9	585.2	1,028	1,121	1,213	1,609*	1,563*	966.9	1,352	1,530	2,215*	2,144*
	(-0.196)	(0.613)	(0.983)	(1.493)	(1.477)	(1.348)	(1.384)	(1.558)	(1.716)	(1.689)	(0.959)	(1.254)	(1.564)	(1.772)	(1.762)
Border Countries ==1	1,427***	1,460***	1,558***	1,695***	1,673***	1,937***	1,960***	1,846**	2,020***	2,008***	3,317***	3,363***	3,334***	3,647***	3,610***
	(2.967)	(2.972)	(2.911)	(3.238)	(3.152)	(2.602)	(2.665)	(2.460)	(2.755)	(2.685)	(2.997)	(3.030)	(2.981)	(3.269)	(3.185)
Geographical Distance	-0.0694	-0.0602*	-0.0374	-0.0396	-0.0362	-0.0667	-0.0642	-0.0364	-0.0417	-0.0406	-0.134	-0.122	-0.0716	-0.0798	-0.0746
	(-1.619)	(-1.760)	(-1.307)	(-1.128)	(-1.138)	(-0.956)	(-0.986)	(-0.695)	(-0.691)	(-0.681)	(-1.258)	(-1.348)	(-0.941)	(-0.860)	(-0.875)
Log Reporter Population	-223.2	-223.2	-223.2	-189.4	-189.4	-7,849	-7,849	-7,849	-8,018	-8,018	-8,527	-8,527	-8,527	-8,745	-8,745
	(-0.123)	(-0.123)	(-0.123)	(-0.0994)	(-0.0994)	(-1.294)	(-1.294)	(-1.294)	(-1.285)	(-1.285)	(-1.227)	(-1.227)	(-1.227)	(-1.225)	(-1.225)
Log Partner Population	2,444	2,444	2,444	2,411	2,411	-9,468	-9,468	-9,468	-9,491	-9,491	-6,420	-6,420	-6,420	-6,439	-6,439
	(1.152)	(1.152)	(1.152)	(1.114)	(1.114)	(-1.225)	(-1.225)	(-1.225)	(-1.182)	(-1.182)	(-0.822)	(-0.822)	(-0.822)	(-0.791)	(-0.791)
Constant	-55,180	-56,089	-55,690	-51,069	-51,091	652,707	652,513	651,733	688,130*	688,217*	662,366	661,293	660,835	704,978	705,000
	(-0.523)	(-0.530)	(-0.528)	(-0.466)	(-0.467)	(1.636)	(1.637)	(1.635)	(1.645)	(1.646)	(1.548)	(1.546)	(1.546)	(1.574)	(1.574)
Observations	13,068	13,068	13,068	12,276	12,276	13,068	13,068	13,068	12,276	12,276	13,068	13,068	13,068	12,276	12,276
R-squared	0.139	0.135	0.138	0.138	0.139	0.078	0.078	0.080	0.080	0.080	0.130	0.129	0.133	0.132	0.133

Pooled Cross-Section of all 35 OECD countries for all years 2002-2012 with a time trend with multi-way clustered standard errors on the reporting and partner country. Each column represents a regression in Euros (mn, real). All regressions use a single linguistic variable, not all combined. Controlling for reporter and partner GDP per capita, common legal origin, in Europe dummy, border countries dummy, geographical distance, log of the reporting country's population, and log of the partner country's population. Also include country specific fixed effects dummies for the reporting and partner country. Coefficients of control variables can be provided upon request.

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4. Bilaterally by Year

DEBT	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Common Official Language	213.1	801.4	2,688	3,195	5,417	-251.8	-3,397*	323.3	-1,119	817.5	-2,715
	(0.0968)	(0.376)	(1.076)	(1.248)	(1.599)	(-0.0820)	(-1.647)	(0.183)	(-0.634)	(0.708)	(-1.087)
Common Spoken Language	1,378	2,054	3,042	3,057	6,908	2,550	-2,991	-3,036**	-2,081	-1,558	-3,856
	(0.838)	(0.917)	(1.320)	(1.088)	(1.591)	(0.742)	(-1.583)	(-2.231)	(-0.992)	(-1.481)	(-1.521)
Common Native Language	-929.4	7,996	10,212	10,161	18,157*	7,087	-4,340	925.5	-3,243	103.8	-7,795
	(-0.373)	(1.630)	(1.524)	(1.452)	(1.917)	(0.980)	(-0.919)	(0.287)	(-0.690)	(0.0633)	(-1.150)
Linguistic Proximity (Ethnologue)	2,417	4,009	5,307	5,887*	9,702	5,843	-2,451	-993.8	-2,463	-443.8	-4,381
	(1.360)	(1.109)	(1.220)	(1.745)	(1.609)	(1.062)	(-1.056)	(-0.640)	(-0.991)	(-0.264)	(-1.357)
Linguistic Proximity (ASJP)	1,491	4,316	5,990	6,447	11,497	5,465	-2,826	-1,122	-2,433	-147.6	-4,491
	(0.852)	(1.150)	(1.221)	(1.372)	(1.621)	(0.937)	(-1.046)	(-0.685)	(-0.920)	(-0.0900)	(-1.265)

EQUITY	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Common Official Language	4,173*	335.6	42.19	1,649*	1,911	2,903***	2,826*	1,665	2,246	2,238**	205.8
	(1.777)	(0.410)	(0.0458)	(1.739)	(1.133)	(2.797)	(1.706)	(1.624)	(1.602)	(2.233)	(0.184)
Common Spoken Language	2,141	1,611	757.1	1,183	1,276	3,822**	2,422	1,445	2,546	2,422**	2,628**
	(1.367)	(1.497)	(0.733)	(0.713)	(0.885)	(2.112)	(1.467)	(1.147)	(1.347)	(2.042)	(2.252)
Common Native Language	5,899*	622.9	915.5	3,093	2,341	4,957*	3,265	3,179	4,199	6,464***	1,591
	(1.741)	(0.604)	(0.413)	(1.490)	(0.901)	(1.698)	(0.804)	(0.929)	(1.299)	(2.675)	(1.005)
Linguistic Proximity (Ethnologue)	2,957*	1,171	674.2	1,639	1,491	4,398**	1,805	1,875	2,871*	3,053**	1,449*
	(1.778)	(1.352)	(0.610)	(1.096)	(1.072)	(2.338)	(1.088)	(1.061)	(1.675)	(2.200)	(1.648)
Linguistic Proximity (ASJP)	3,064*	1,205	934.7	1,753	1,646	4,484**	2,654	2,194	3,063	4,061**	1,943**
	(1.717)	(1.265)	(0.746)	(1.089)	(1.159)	(2.171)	(1.087)	(1.029)	(1.500)	(2.379)	(2.009)

VALUE	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Common Official Language	2,868	1,137	2,730	4,844*	7,328*	2,652	-571.0	1,988	1,127	3,055*	-2,509
	(0.904)	(0.537)	(0.973)	(1.655)	(1.868)	(0.792)	(-0.212)	(1.094)	(0.518)	(1.879)	(-0.957)
Common Spoken Language	2,404	3,665	3,800	4,240	8,184*	6,372	-568.8	-1,592	464.8	863.5	-1,228
	(0.923)	(1.132)	(1.425)	(1.043)	(1.725)	(1.349)	(-0.321)	(-1.095)	(0.180)	(0.512)	(-0.551)
Common Native Language	5,061	8,619	11,128	13,254	20,498*	12,044	-1,075	4,104	955.6	6,568*	-6,204
	(1.179)	(1.545)	(1.479)	(1.496)	(1.889)	(1.375)	(-0.180)	(1.043)	(0.164)	(1.898)	(-0.987)
Linguistic Proximity (Ethnologue)	4,572	5,180	5,981	7,525	11,192*	10,241	-646.1	881.2	408.1	2,610	-2,932
	(1.527)	(1.226)	(1.404)	(1.536)	(1.660)	(1.577)	(-0.250)	(0.412)	(0.137)	(0.944)	(-0.932)
Linguistic Proximity (ASJP)	4,334	5,520	6,925	8,200	13,142*	9,949	-171.7	1,072	630.2	3,914	-2,548
	(1.456)	(1.220)	(1.361)	(1.273)	(1.670)	(1.439)	(-0.0549)	(0.457)	(0.187)	(1.307)	(-0.733)

Bilaterally by year for all 35 OECD countries for all years 2002-2012 with multi-way clustered standard errors on the reporting and partner country. Each cell (and its corresponding standard error) represents a regression in Euros (mn, real). All regressions use a single linguistic variable, not all combined, for an individual year. Controlling for reporter and partner GDP per capita, common legal origin, in Europe dummy, border countries dummy, geographical distance, log of the reporting country's population, and log of the partner country's population. Also include country specific fixed effects dummies for the reporting and partner country. Coefficients of control variables can be provided upon request.

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5. *Pooled Pre-2008 with Time Trend*

VARIABLES	Equity					Debt					Value				
Common Official Language	1,836**					2,010					3,593				
	(2.089)					(1.066)					(1.629)				
Common Spoken Language		1,798*					3,165					4,777*			
		(1.758)					(1.402)					(1.664)			
Common Native Language			2,971*					8,781					11,767*		
			(1.829)					(1.608)					(1.764)		
Linguistic Proximity (Ethnologue)				2,055**					5,527*					7,449*	
				(2.070)					(1.658)					(1.828)	
Linguistic Proximity (ASJP)					2,181**					5,868					8,012*
					(2.008)					(1.521)					(1.699)
Year Time Trend	114.2	114.2	114.2	125.8	125.8	204.5	204.5	204.5	220.7	220.7	142.1	142.1	142.1	163.6	163.6
	(1.267)	(1.267)	(1.267)	(1.313)	(1.313)	(1.437)	(1.437)	(1.437)	(1.435)	(1.435)	(0.852)	(0.852)	(0.852)	(0.897)	(0.897)
Reporter GDP per capita	0.00159	0.00159	0.00159	0.00146	0.00146	-0.0228	-0.0228	-0.0228	-0.0252	-0.0252	-0.0141	-0.0141	-0.0141	-0.0169	-0.0169
	(0.0989)	(0.0975)	(0.0973)	(0.0802)	(0.0809)	(-0.463)	(-0.459)	(-0.465)	(-0.496)	(-0.497)	(-0.252)	(-0.234)	(-0.245)	(-0.213)	(-0.278)
Partner GDP per capita	-0.00765	-0.00765	-0.00765	-0.00801	-0.00801	0.0512	0.0512	0.0512	0.0537	0.0537	0.113	0.113	0.113	0.118	0.118
	(-0.298)	(-0.302)	(-0.295)	(-0.308)	(-0.307)	(1.370)	(1.319)	(1.357)	(1.390)	(1.370)	(1.575)	(1.562)	(1.558)	(1.508)	(1.568)
Common Legal Origin	67.84	220.5	111.7	82.69	11.60	858.4	880.3	297.8	446.6	255.2	943.2	1,089	373.7	520.1	244.2
	(0.226)	(0.687)	(0.510)	(0.280)	(0.0445)	(1.302)	(1.054)	(0.694)	(0.742)	(0.489)	(1.105)	(1.013)	(0.660)	(0.657)	(0.357)
In Europe ==1	-323.5	-8.142	68.55	311.9	263.0	-1,281	-933.8	-704.2	-751.5	-883.0	-1,574	-955.4	-648.4	-461.0	-628.8
	(-0.678)	(-0.0187)	(0.180)	(0.612)	(0.522)	(-0.898)	(-0.806)	(-0.623)	(-0.643)	(-0.727)	(-0.931)	(-0.702)	(-0.508)	(-0.357)	(-0.472)
Border Countries ==1	1,451**	1,514**	1,600**	1,724**	1,714**	1,962	1,871	1,839	2,093	2,066	3,326*	3,280*	3,310*	3,691**	3,648**
	(2.343)	(2.181)	(2.173)	(2.439)	(2.371)	(1.409)	(1.342)	(1.277)	(1.484)	(1.434)	(1.952)	(1.943)	(1.890)	(2.212)	(2.133)
Geographical Distance	-0.0711*	-0.0625*	-0.0421	-0.0422	-0.0416	-0.158	-0.145	-0.0799	-0.103	-0.101	-0.224	-0.204	-0.118	-0.142	-0.139
	(-1.804)	(-1.794)	(-1.264)	(-1.170)	(-1.193)	(-1.364)	(-1.366)	(-0.892)	(-1.160)	(-1.138)	(-1.515)	(-1.548)	(-1.035)	(-0.337)	(-1.223)
Log Reporter Population	3,942	3,942	3,942	4,190	4,190	3,568	3,568	3,568	2,999	2,999	5,410	5,410	5,410	4,824	4,824
	(0.952)	(0.952)	(0.952)	(0.938)	(0.938)	(0.298)	(0.298)	(0.298)	(0.244)	(0.244)	(0.345)	(0.345)	(0.345)	(0.296)	(0.296)
Log Partner Population	137.5	137.5	137.5	-384.1	-384.1	6,886	6,886	6,886	6,794	6,794	6,070	6,070	6,070	5,457	5,457
	(0.0317)	(0.0317)	(0.0317)	(-0.0859)	(-0.0859)	(0.830)	(0.830)	(0.830)	(0.801)	(0.801)	(0.575)	(0.575)	(0.575)	(0.508)	(0.508)
Constant	-296,311	-297,055	-296,707	-315,785	-315,673	-584,253	-585,949	-586,504	-608,843	-608,544	-477,500	-479,883	-480,232	-504,601	-504,271
	(-1.483)	(-1.486)	(-1.486)	(-1.510)	(-1.510)	(-1.464)	(-1.466)	(-1.467)	(-1.455)	(-1.454)	(-1.006)	(-1.010)	(-1.010)	(-1.002)	(-1.001)
Observations	7,128	7,128	7,128	6,696	6,696	7,128	7,128	7,128	6,696	6,696	7,128	7,128	7,128	6,696	6,696
R-squared	0.157	0.153	0.155	0.157	0.157	0.214	0.213	0.221	0.219	0.220	0.249	0.248	0.256	0.254	0.256

Pooled cross-section of all 35 OECD countries for the years 2002-2007 with a time trend and multi-way clustered standard errors on the reporting and partner country. Each column represents a regression in Euros (mn, real). All regressions use a single linguistic variable, not all combined. Controlling for reporter and partner GDP per capita, common legal origin, in Europe dummy, border countries dummy, geographical distance, log of the reporting country's population, and log of the partner country's population. Also include country specific fixed effects dummies for the reporting and partner country. Coefficients of control variables can be provided upon request.

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6. Pooled Post-2008 with Time Trend

VARIABLES	Equity					Debt					Value				
Common Official Language	1,836***					-1,218					618.1				
	(2.612)					(-1.427)					(0.592)				
Common Spoken Language		2,292**					-2,705**					-412.2			
		(2.019)					(-2.188)					(-0.317)			
Common Native Language			3,740**					-2,870					869.8		
			(2.399)					(-1.408)					(0.351)		
Linguistic Proximity (Ethnologue)				2,211*					-2,147*					64.13	
				(1.852)					(-1.703)					(0.0490)	
Linguistic Proximity (ASJP)					2,783**					-2,204					579.3
					(2.039)					(-1.623)					(0.376)
Year Time Trend	-43.55	-43.55	-43.55	-48.82	-48.82	29.67	29.67	29.67	30.95	30.95	-13.87	-13.87	-13.87	-17.87	-17.87
	(-0.626)	(-0.626)	(-0.626)	(-0.679)	(-0.679)	(0.229)	(0.229)	(0.229)	(0.226)	(0.226)	(-0.0800)	(-0.0800)	(-0.0800)	(-0.0978)	(-0.0978)
Reporter GDP per capita	-0.0282*	-0.0282*	-0.0282*	-0.0287*	-0.0287*	-0.0441	-0.0441	-0.0441	-0.0448	-0.0448	-0.0723	-0.0723	-0.0723	-0.0735	-0.0735
	(-1.832)	(-1.781)	(-1.846)	(-1.901)	(-1.877)	(-0.882)	(-0.871)	(-0.881)	(-0.851)	(-0.853)	(-1.152)	(-1.152)	(-1.159)	(-1.118)	(-1.125)
Partner GDP per capita	-0.0740	-0.0740	-0.0740	-0.0766	-0.0766	0.0113	0.0113	0.0113	0.0127	0.0127	-0.0627	-0.0627	-0.0627	-0.0640	-0.0640
	(-1.260)	(-1.258)	(-1.257)	(-1.262)	(-1.262)	(0.210)	(0.209)	(0.210)	(0.223)	(0.224)	(-0.809)	(-0.814)	(-0.815)	(-0.798)	(-0.800)
Common Legal Origin	135.5	228.1	94.85	128.0	-23.52	-36.44	46.01	33.38	-75.50	-14.01	99.07	274.1	128.2	52.46	-37.53
	(0.540)	(0.788)	(0.483)	(0.540)	(-0.105)	(-0.0800)	(0.101)	(0.0723)	(-0.152)	(-0.0284)	(0.192)	(0.481)	(0.245)	(0.0930)	(-0.0666)
In Europe ==1	218.7	534.8	631.3	985.0	971.9	3,798*	3,587*	3,514	4,441	4,498	4,017	4,121	4,145*	5,426*	5,470*
	(0.258)	(0.614)	(0.796)	(1.018)	(1.008)	(1.674)	(1.649)	(1.566)	(1.621)	(1.629)	(1.589)	(1.638)	(1.657)	(1.697)	(1.713)
Border Countries ==1	1,399**	1,396**	1,508**	1,661**	1,625**	1,907*	2,067*	1,854*	1,933*	1,940*	3,306**	3,463**	3,363**	3,594**	3,564**
	(2.055)	(2.112)	(2.124)	(2.270)	(2.207)	(1.690)	(1.774)	(1.687)	(1.657)	(1.671)	(2.166)	(2.239)	(2.208)	(2.252)	(2.234)
Geographical Distance	-0.0673	-0.0574	-0.0318	-0.0365	-0.0298	0.0426	0.0332	0.0158	0.0315	0.0320	-0.0247	-0.0242	-0.0160	-0.00500	0.00221
	(-1.150)	(-1.216)	(-0.830)	(-0.766)	(-0.664)	(0.453)	(0.326)	(0.159)	(0.245)	(0.253)	(-0.180)	(-0.189)	(-0.135)	(-0.0312)	(0.0142)
Log Reporter Population	6,586	6,586	6,586	6,539	6,539	30,088	30,088	30,088	30,694	30,694	36,674	36,674	36,674	37,233	37,233
	(1.475)	(1.475)	(1.475)	(1.540)	(1.540)	(1.372)	(1.372)	(1.372)	(1.333)	(1.333)	(1.337)	(1.337)	(1.337)	(1.302)	(1.302)
Log Partner Population	18,118	18,118	18,118	18,400	18,400	7,012	7,012	7,012	6,613	6,613	25,131	25,131	25,131	25,013	25,013
	(1.087)	(1.087)	(1.087)	(1.075)	(1.075)	(0.457)	(0.457)	(0.457)	(0.404)	(0.404)	(0.933)	(0.933)	(0.933)	(0.890)	(0.890)
Constant	-322,423	-323,531	-323,072	-316,445	-316,626	-681,886*	-680,278*	-681,327*	-686,454*	-686,622*	-1.004e+06*	-1.004e+06*	-1.004e+06*	-1.003e+06*	-1.003e+06*
	(-1.122)	(-1.125)	(-1.124)	(-1.079)	(-1.080)	(-1.932)	(-1.929)	(-1.930)	(-1.896)	(-1.897)	(-1.830)	(-1.829)	(-1.830)	(-1.790)	(-1.790)
Observations	5,940	5,940	5,940	5,580	5,580	5,940	5,940	5,940	5,580	5,580	5,940	5,940	5,940	5,580	5,580
R-squared	0.142	0.139	0.143	0.141	0.143	0.032	0.032	0.032	0.034	0.034	0.066	0.066	0.066	0.069	0.069

Pooled cross-section of all 35 OECD countries for the years 2008-2012 with a time trend and multi-way clustered standard errors on the reporting and partner country. Each column represents a regression in Euros (mn, real). All regressions use a single linguistic variable, not all combined. Controlling for reporter and partner GDP per capita, common legal origin, in Europe dummy, border countries dummy, geographical distance, log of the reporting country's population, and log of the partner country's population. Also include country specific fixed effects dummies for the reporting and partner country. Coefficients of control variables can be provided upon request.

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1